January 2020

HIGH RESOLUTION EYE TRACKING CAMERA ENABLING A FOVEATED IMAGING FOR VIDEO COLLABORATION DEVICES

Farhad Abbassi Gholmansaraei

HIGH RESOLUTION EYE TRACKING CAMERA ENABLING A FOVEATED IMAGING FOR VIDEO COLLABORATION DEVICES

AUTHORS:

Farhad Abbassi Gholmansaraei

ABSTRACT

Presented herein are foveated imaging techniques with novel video camera image sensor readout mechanism that is suitable for use with high resolution image sensors. The techniques presented herein enable rapid/fast frame rate video capture and camera operation in both "Tele mode" and "Wide mode." Also presented herein is the use of a single high resolution camera (e.g., resolution greater than 100MP) for video conferencing/collaboration, implementations for foveated imaging, implementations for equivalent of optical zooming for Tele mode, and delivery of a fast video capture in Tele mode with full resolution.

DETAILED DESCRIPTION

Camera array technology with optical zoom and fusion has been introduced in recent years on mobile devices (e.g., mobile phones, tablet computers, etc.). These cameras are fusing the images of two cameras (within a camera array) to achieve optical zoom, high resolution, low light performance, etc. Due to the rendering workload, the image fusion is mainly applied to still images or videos with low frame rate.

Certain vendors are also integrating camera array technology into endpoint collaboration devices and moving away from mechanical pan tilt zoom (PTZ) cameras. In fact, recent product offering from some vendors include camera arrays with both Wide and Tele lenses. For example, Figures 1A and 1B, below illustrate an example product with one Wide camera and three (3) Tele cameras, where the combination of the Tele and Wide cameras provide optical zoom functionality. Figure 1A is a perspective view of the product, while Figure 1B is schematic diagram illustrating the one Wide camera and three Tele cameras.

Figure 1A

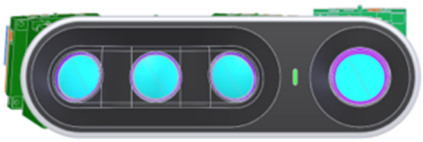

Figure 1B

In above configuration, the field of view (FOV) of the single Wide camera covers the entire conference room, while the Tele cameras are used to focus on the speakers (or actions) during the conference call and provide a sharper image than the Wide camera. The data to be transmitted (e.g., Wide camera capture or any one of Tele camera's captures) from the endpoint collaboration devices is determined by the image processing or computer vision programs. This would be the image and view displayed in remote user's display.

The Rolling Shutter CMOS image sensor resolutions are now exceeding 100MP and it is projected that the resolution approaches to ~0.5GP (500MP) in few years. This will open the opportunities to exploit the benefits of Gigapixel imaging. Gigapixel imaging has demonstrated the benefits of using high-resolution imaging and being able to do an effective digital zooming and abort any optical zooming[1]. Figure 2, below, depicts the high-quality digital zooming that is possible with Giga-pixel imaging[1]. With GigaPixel imaging, the "digital zoom" image sharpness can be as good as the "Optical Zoom."

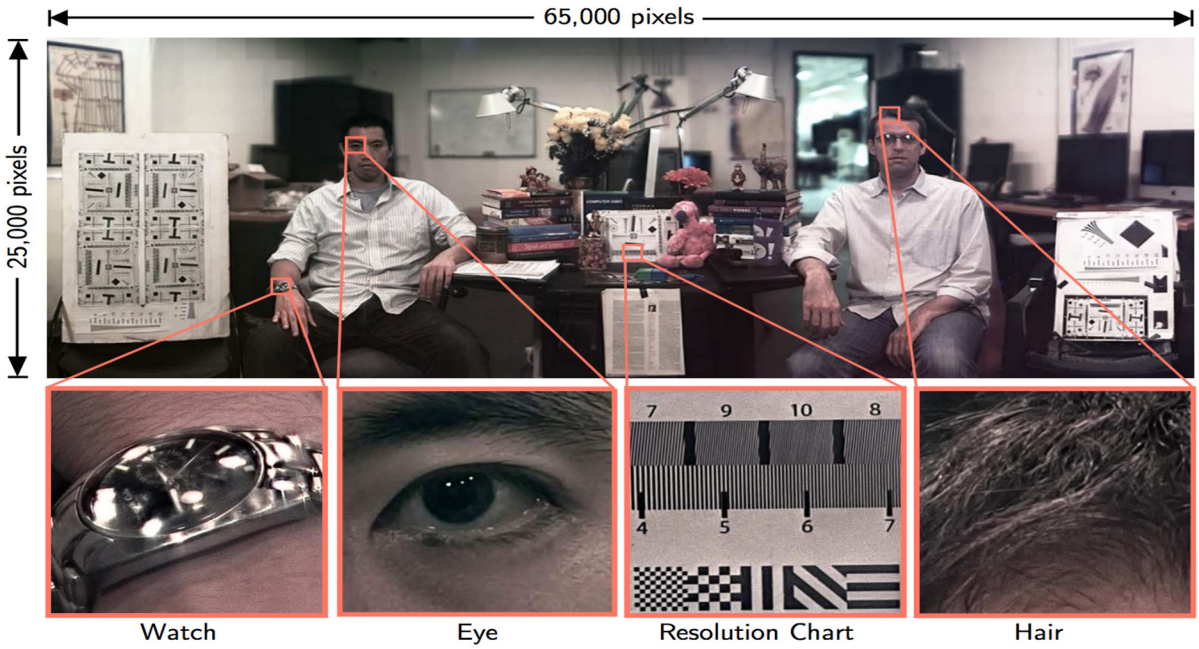

Figure 2[1]

The techniques presented herein propose the use of a single camera with a high resolution image sensor (e.g., resolution greater than 100MP) to replace conventional camera arrays with optical zoom capabilities. Since the high resolution image sensors have much smaller pixel pitch and higher pixel counts, the quality of digital zooming would be equivalent to the quality of shorter optical zoom that is currently utilized in mobile devices or video conferencing products. This means that single camera with high resolution image sensor and a foveated imaging technique can be utilized to operate as a Wide and/or Tele camera, while delivering the same video quality that is achievable with camera arrays. The selection of the Wide or Tele mode is enabled by artificial intelligence (AI) vision processing. However, the major obstacle in use of high resolution image sensor (greater than 100MP) for video applications is the slower frame rate for high resolution mode (when all pixels are readout and there is no binning). Presented herein are techniques to overcome this issue by using a novel readout mode that is enabled by AI vision processing.

A single camera equipped with a high resolution image sensor and a lens with a wide field of view (FOV) can emulate an eye tracking system. Figure 3, below, illustrates a block diagram of an eye tracking camera with foveated imaging.

.

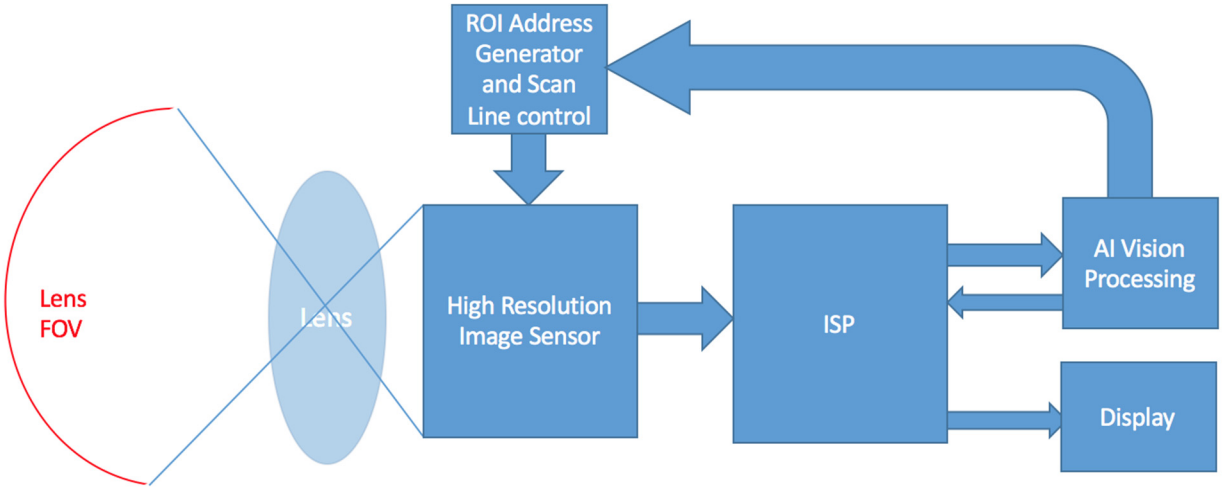

Figure 3

In Figure 3, the FOV of the lens (lens FOV) exposes the high resolution image sensor. The image signal processing block performs the pipeline correction and other image signal processing tasks that are common for all color cameras. The image that is processed by the image signal processing will feed the AI vision processing block that monitors all activities within the Lens FOV, such as tracking of people, objects, defined events, interactions with humans, and can determine whether the camera to operates in Wide mode or Tele. The AI vision processing block is the brain of the system that controls the image sensor readout and the image signal processing output image to display.

In Wide mode, the image sensor is readout in the same manner as any rolling shutter image sensor and the image signal processing transmits the entire image sensor image to display. Pixel binning is used in Wide mode to increase the frame rate that would result in lower image resolution.

In Tele mode, the AI block detects an 'event' within the lens FOV, determines the location of the 'event' within the image, generates a region of interest (ROI) address associated with the event, and initiates the Tele mode which zoom (digital zoom) on the specific event. For Video conferencing applications, the 'event' could be speaker tracking or presenter tracking in which the Tele mode would be activated. There will be two regions in Tele modes; (a) ROI (Region of interest), and (b) ROA (region of awareness). The ROA is the entire image, including the region outside of the ROI. In Tele Mode, the image signal

processing only transmits the ROI to the display. However, the AI block processes and evaluates the entire image (ROA) to determine the next 'event' or 'events' in order to generate a new ROI or switch back to the Wide mode.

In Tele mode, the image sensor uses a special and novel scanning mode that has a different readout pattern for ROI compared to rest of the image sensor region. ROI is a smaller region (compared to the entire image sensor area). The AI processing block provides the address of the ROI and controls how to readout the image sensor row lines. The 'address generator and Row line scan control' oversees the image sensor readout during the Tele mode as seen in Figure 4, below. That is, the ROI address and Row scan control block governs the image sensor readout during Tele mode

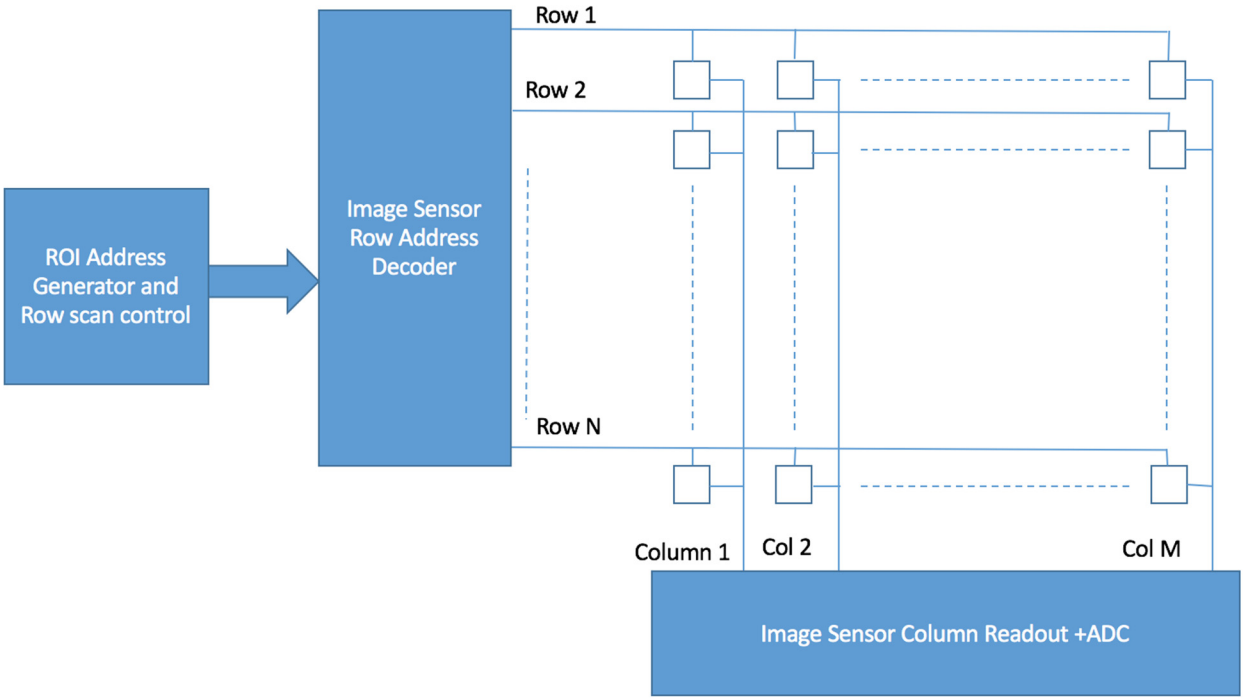

Figure 4

In Tele mode, the camera should operate in a high-resolution mode in order to deliver a high quality digital zoom. Since no pixel binning is allowed during the high resolution mode, and the image sensor has to read all individual pixels, the frame rate will

be lower. To overcome this issue, proposed herein is a new scanning mode, which is illustrated in Figure 5, below.

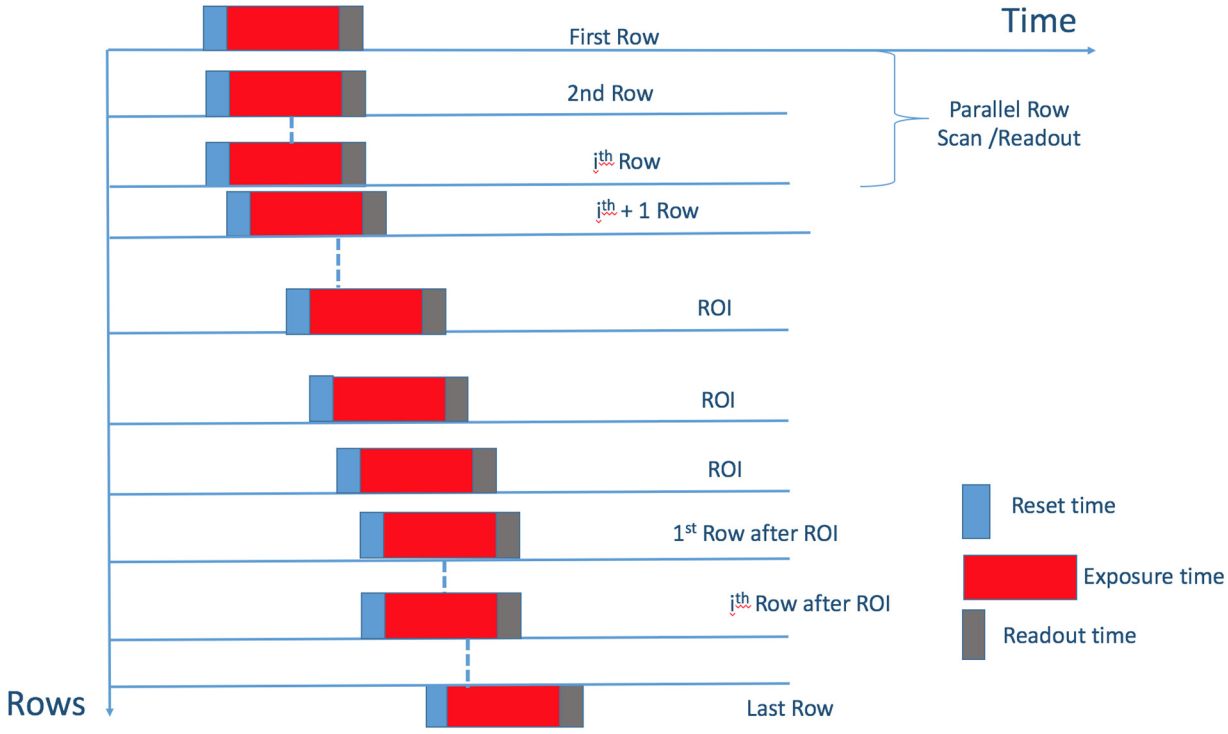

Figure 5

In Figure 5 (special mode for Tele ROI scanning), there are two regions in Tele mode, namely the region of interest (ROI) and the region outside of the ROI (non-ROI regions). The rows in the non-ROI regions are read simultaneously and in parallel in order to reduce the frame readout time and increase video frame rate. As shown in Figure 5, there are 'i' rows that are readout concurrently (same reset, exposure, and readout). The 'i' is determined by the AI vision processing block. Within the ROI, each row is scanned individually and therefore, the ROI image resolution is higher than the regions outside of ROI. The two factors that the AI block analyzes to determine the 'i' are the 'required frame rate for Tele mode' and the minimum image resolution that is required for ROA to detect the next event (new ROI for subsequent frames if needed).

The above readout mechanism will assure that the ROI region is readout with the highest resolution (high resolution Tele mode) without sacrificing the video frame rate (that is typical for standard rolling shutter image sensors).

In one example video conferencing application in accordance with the techniques presented herein, it is assumed that a CMOS Image sensor with resolution of 360MP and pixel size of 0.5µm is provided. In Wide mode, the camera output (to display) will cover the entire conference room that is within the lens Field of View (FOV). Also in Wide mode, 3x3 image binning will be used, which results in an image resolution of 40MP (pixel size of 1.5µm). The frame rate for 40MP image will be 60fps. It this example, an event occurs and the camera would like to zoom on a speaker or presenter. The camera will be switched to Tele mode and will output only the ROI to display. The ROI region in the Tele mode will be read with full image resolution (no binning), which will have the pixel size of 0.5µm. This results in a zoom that is equivalent to an optical zoom of three (3) (1.5/0.5 = 3). Therefore, in Tele mode, a zoom that is equivalent to a x3 optical zoom is provided. These techniques also overcome the slower video frame rate for Tele mode by using the readout scheme shown in Figure 5 (e.g., using ”i = 10” and reading 10 row line simultaneously) to achieve a frame rate of 60 for Tele mode.

As noted, the techniques described above outline a foveated imaging technique and a novel video camera’s image sensor readout mechanism that is suitable for high resolution image sensors to enable fast frame rate video capture. Also as noted, these techniques are based on a single high resolution camera (e.g., resolution greater than 100MP) that would switch between Wide and Tele modes.

- At Wide FOV, the camera is used as eye tracker (e.g., in a video conferencing room). To achieve a fast frame rate (like 60fps), the image sensor is run in binning mode (low resolution mode) during Wide mode.
- An AI/ML-powered vision processing block understands the environment and recognizes objects, people, and human within the eye tracker FOV. AI vision processing determines when to switch from Wide to Tele mode (enable the foveated imaging), from one Tele position to another, or switch back from Tele to Wide.
- The camera’s Tele mode is a high resolution mode and is controlled by the AI-vision processing output to track objects, people or interact with humans. In Tele mode, the image sensor is not using pixel binning and therefore, delivering its highest resolution (smallest pixel size).

- The AI vision processing block defines the region of interest (ROI) region by generating an address associated with ROI. The output of the address generator is connected to the row control of the image sensor.
- In Tele mode, the image sensor uses a special scanning mode that only effectively scans the ROI selected by the AI-powered vision processing.
- The image sensor of the camera in Tele mode:
    - Scans the ROI region in full resolution mode (to achieve highest resolution for ROI). This means that all rows of ROI are scanned individually.
    - Scans the rows outside of ROI in low resolution mode by selecting / reading 'i' rows in parallel. This enhances the frame readout speed for Tele mode and also enables the AI-powered vision processing to check the activities on entire image for the next ROI selection (for Tele) or for switching back to Wide mode.
    - The exposure time can be programmed to any time for a best compromise of signal-to-noise ratio (SNR) and faster frame rate.
    - The camera image signal processing only displays the ROI area in Tele mode by cropping the capture image sensor.
- Image fusion and foveated imaging of Tele camera (controlled by AI) view with Wide camera view to achieve enhanced video picture quality and equivalent optical zooming for Tele mode.

References:

1. Oliver S. Cossairt, Daniel Miau, and Shree K. Nayar "A Scaling Law for Computational Imaging Using Spherical Optics", **Journal of the Optical Society of America A**, Vol. 28, Issue 12, pp. 2540-2553, (2011), *www1.cs.columbia.edu › CAVE › publications › pdfs › Cossairt_JOSA11.*